

HD-A131 567

SELECTION PROCEDURES FOR OPTIMAL SUBSETS OF REGRESSION
VARIABLES(U) PURDUE UNIV LAFAYETTE IN DEPT OF
STATISTICS S S GUPTA ET AL. JUL 83 TR-83-29

1/1

UNCLASSIFIED

N00014-75-C-0455

F/G 12/1

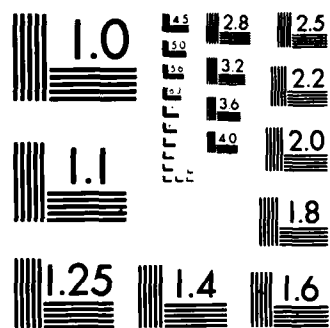
NL

END

FILED

10

10/10/83

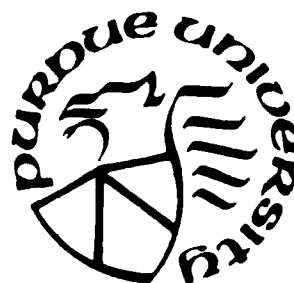


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A 131567

12

PURDUE UNIVERSITY



DEPARTMENT OF STATISTICS

DTIC
ELECTE
S AUG 19 1983

D

DTIC FILE COPY

88 08 17 008

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

SELECTION PROCEDURES FOR OPTIMAL SUBSETS OF
REGRESSION VARIABLES*¹

by

Shanti S. Gupta
Purdue University
D. Y. Huang and C. L. Chang
National Taiwan Normal University
Taipei, Taiwan, ROC

Technical Report #83-29

Department of Statistics
Purdue University

July 1983

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



*¹ This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

SELECTION PROCEDURES FOR OPTIMAL SUBSETS OF REGRESSION VARIABLES*¹

by

Shanti S. Gupta
Purdue University

D. Y. Huang and C. L. Chang
National Taiwan Normal University
Taipei, Taiwan, ROC

I. INTRODUCTION

In the recent years, a number of authors have studied the problem of selecting the "best" or a "good" subset of regression variables in the format of selection and ranking theory. Arvesen and McCabe [1] considered a procedure for selecting the best model from among all reduced models involving r (fixed) out of p independent variables. Huang and Panchapakesan [4] discussed the problem of eliminating all inferior models using the criterion of expected residual sum of squares to define inferior models. Hsu and Huang [3] investigated a sequential procedure for selecting good regression models.

Consider the usual linear model

$$(1.1) \quad \underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

where $X = [1, \underline{x}_1, \dots, \underline{x}_{p-1}]$ is an $n \times p$ matrix of known constants,

$\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})$ is a $1 \times p$ vector of unknown parameters and

$\underline{\epsilon} \sim N(\underline{0}, \sigma_0^2 I_n)$. Here I_n denotes the identity matrix of order $n \times n$ and

$\underline{a} = (a, a, \dots, a)$. The model (1.1) having $p-1$ independent variables is

*¹This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

considered as the true model. Any reduced model whose "X matrix" has r columns is obtained by retaining any $r-1$ of the $p-1$ independent variables, where $2 \leq r \leq p-2$. There are $k_r = \binom{p-1}{r-1}$ such models. These k_r reduced models of "size" r are indexed arbitrarily by i going from 1 to k_r . We will refer to a typical model in this consider as Model ri . A reduced model of size r can be written as

$$(1.2) \quad \underline{Y} = X_{ri}\underline{\beta}_{ri} + \underline{\epsilon}_{ri}, \quad i = 1, 2, \dots, k_r,$$

where X_{ri} and $\underline{\beta}_{ri}$ are obtained from X and $\underline{\beta}$ corresponding to the variables that are retained in the model, and $\underline{\epsilon}_{ri} \sim N(\underline{0}, \sigma_{ri}^2 I_n)$.

It should be first noted that our comparisons of models are made under the true model assumptions. Any reduced model with the associated error variance σ_{ri}^2 is called inferior if $\sigma_{ri}^2 \geq \Delta \sigma_0^2$ where $\Delta (> 1)$ is a specified constant. Our goal is to eliminate all inferior models, from the set of $2^{p-1}-1$ regression models including the true model. For this purpose, we consider a family of hypotheses testing problems, namely, $H_{0,ri}$ against K_{ri} , $i = 1, \dots, k_r$; $r = 2, \dots, p$. Rejecting $H_{0,ri}$ will mean declaring Model ri to be inferior. We derive our rule in Section III subject to controlling errors as explained therein.

II. PRELIMINARIES

For any r , $2 \leq r \leq p$, we know that

$$(2.1) \quad \begin{aligned} SS_{ri} &= \underline{Y}' \{I - X_{ri}(X_{ri}'X_{ri})^{-1}X_{ri}'\} \underline{Y} \\ &= \underline{Y}' Q_{ri} \underline{Y}, \text{ say,} \end{aligned}$$

and

$$(2.2) \quad \frac{SS_{ri}}{\sigma_0^2} \sim \chi^2_{\{v_r, \lambda_{ri}\}} \quad \text{under the true model,}$$

where the degrees of freedom $\nu_r = n-r$, and the noncentrality parameter $\lambda_{ri} = (\underline{X}_\beta)' Q_{ri}(\underline{X}_\beta)/2$, $1 \leq i \leq k_r$. We note that λ_{ri} is not, in general, zero and

$$(2.3) \quad \sigma_{ri}^2 = \sigma_0^2 + \frac{2}{\nu_r} \sigma_0^2 \lambda_{ri}.$$

It is clear from (2.3) that λ_{ri} should not be large for a good model.

For convenience, we use $\tilde{\beta}_{ri}$ to denote the vector obtained from $\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})$ replacing with zeros the β_j associated with variables that are dropped from the full model. For examples, when $p = 6$, $r = 4$ and the variables that are retained are X_1 , X_3 and X_4 , we have $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$, $\beta_{ri} = (\beta_0, \beta_1, \beta_3, \beta_4)$, and $\tilde{\beta}_{ri} = (\beta_0, \beta_1, 0, \beta_3, \beta_4, 0)$.

Finally, we let

$$(2.4) \quad \Omega_{0,ri} = \{\tilde{\beta}_{ri} | \lambda_{ri} = 0\}$$

and

$$(2.5) \quad \Omega_{1,ri} = \{\tilde{\beta}_{ri} | \lambda_{ri} \geq \lambda_r\},$$

where $i = 1, 2, \dots, k_r$; $r = 2, \dots, p$, where $\lambda_r = \nu_r(\Delta-1)/2$.

III. DERIVATION OF THE RULE

As we explained in Section I, our rule is based on a family of hypothesis testing problems. Let $\Omega_1 = \bigcup_{r=2}^p \bigcup_{i=1}^{k_r} \Omega_{1,ri}$, and $\Omega_0 = \bigcap_{r=2}^p \bigcap_{i=1}^{k_r} \Omega_{0,ri}$.

Consider the following family of hypotheses testing problem.

$$(3.1) \quad H_{0,ri}: \tilde{\beta}_{ri} \in \Omega_0 \text{ vs } K_{ri}: \tilde{\beta}_{ri} \in \Omega_{1,ri};$$

$i = 1, \dots, k_r$, $r = 2, \dots, p$. Let φ_{ri} be the test function for $H_{0,ri}$ vs K_{ri} .

Then the simultaneous test of all the hypotheses in (3.1) is defined by the vector φ whose components are φ_{ri} , $i = 1, \dots, k_r$; $r = 2, \dots, p$. The power function of the test is vector of the power functions $p_{ri}(\tilde{g}_{ri})$ of the individual tests, where

$$p_{ri}(\tilde{g}_{ri}) = E_{\tilde{g}_{ri}} \varphi_{ri}(Y),$$

$$i = 1, \dots, k_r; r = 2, \dots, p.$$

Let $S(\alpha)$ be the set of all tests φ_{ri} , $i = 1, \dots, k_r$; $r = 2, \dots, p$ such that

$$(3.2) \quad E_{\tilde{g}_{ri}} \varphi_{ri}(Y) \leq \alpha, \quad \tilde{g}_{ri} \in \Omega_0,$$

where α is the specified value, $i = 1, \dots, k_r$; $r = 2, \dots, p$.

Case 1: When we estimate σ_0^2 and use it as the known value of σ_0^2 , i.e., we treat σ_0^2 as known.

Since

$$v_r S_{ri} = \frac{SS_{ri}}{\sigma_0^2},$$

is distributed with noncentral chi-square $\chi^2(v_r, \lambda_{ri})$, we denote the probability density of S_{ri} as $g_{\lambda_{ri}}(s_{ri})$, $i = 1, \dots, k_r$; $r = 2, \dots, p$. For any r and i , we define

$$\varphi_{ri}^0(Y) = \begin{cases} 1, & \text{if } g_{\lambda_r}(s_{ri}) \geq cg_0(s_{ri}), \\ 0, & \text{if } g_{\lambda_r}(s_{ri}) < cg_0(s_{ri}), \end{cases}$$

such that $E_{\tilde{g}_{ri}} \varphi_{ri}^0(Y) = \alpha$, $\tilde{g}_{ri} \in \Omega_0$, where s_{ri} is the observed value of S_{ri} and $g_0(s_{ri})$ is the central chi-square probability density. It can be shown that φ_{ri}^0 , $i = 1, \dots, k_r$; $r = 2, \dots, p$, maximize

$$\min_{\substack{i=1,\dots,k_r \\ r=2,\dots,p}} \inf_{\tilde{\beta} \in \Omega_{1,ri}} E_{\tilde{\beta}_{ri}} \varphi_{ri}(\underline{y})$$

among all tests φ_{ri} in $S(\alpha)$, $i = 1, \dots, k_r$; $r = 2, \dots, p$.

Since

$$g_{\lambda_{ri}}(s_{ri}) = v_r e^{-\lambda_{ri}} \sum_{\ell=0}^{\infty} \frac{\lambda_{ri}^{\ell} (v_r s_{ri})^{\frac{1}{2}v_r + \ell - 1} e^{-\frac{1}{2}(v_r s_{ri})}}{\ell! 2^{\frac{1}{2}v_r + \ell} \Gamma(\frac{1}{2}v_r + \ell)},$$

for $\lambda_{ri} > 0$, and

$$g_0(s_{ri}) = \frac{v_r (v_r s_{ri})^{\frac{1}{2}v_r - 1}}{\Gamma(\frac{v_r}{2}) 2^{\frac{1}{2}v_r}} e^{-\frac{1}{2}v_r s_{ri}},$$

hence

$$\frac{g_{\lambda_{ri}}(s_{ri})}{g_0(s_{ri})} = \sum_{\ell=0}^{\infty} \frac{e^{-\lambda_{ri}} \lambda_{ri}^{\ell}}{\ell!} \left(\frac{v_r s_{ri}}{2}\right)^{\ell} \frac{\Gamma(\frac{1}{2}v_r)}{\Gamma(\frac{1}{2}v_r + \ell)}$$

is strictly increasing function of s_{ri} .

The rule φ_{ri}^0 is equivalent to the following

$$\varphi_{ri}^0(\underline{y}) = \begin{cases} 1, & \text{if } s_{ri} \geq c_{ri}, \\ 0, & \text{if } s_{ri} < c_{ri}, \end{cases}$$

where c_{ri} is determined by

$$P\{S_{ri} \geq c_{ri} | \lambda_{ri} = 0\} = \alpha.$$

Case 2: When σ_0^2 is unknown.

Since

$$V_{ri} = \frac{(SS_{ri} - SS_{p1})/(p-r)}{SS_{p1}/(n-p)}$$

is distributed with noncentral F distribution $F(p-r, n-p; \lambda_{ri})$ with noncentral parameter λ_{ri} , we denote the probability density of V_{ri} as $f_{\lambda_{ri}}(v_{ri})$, and

$$\frac{f_{\lambda_{ri}}(v_{ri})}{f_0(v_{ri})} = \sum_{j=0}^{\infty} \frac{\lambda_{ri}^j e^{-\lambda_{ri}} \Gamma(\frac{2j+n-r}{2}) \Gamma(\frac{n-r}{2})}{j! \Gamma(\frac{2j+p-r}{2}) \Gamma(\frac{n-r}{2})} \left(\frac{p-r}{n-p}\right)^j \left(\frac{v_{ri}}{1 + \frac{p-r}{n-p} v_{ri}}\right)^j$$

is strictly increasing in v_{ri} , for $i = 1, \dots, k_r$; $r = 2, \dots, p$.

We define

$$\psi_{ri}^0(y) = \begin{cases} 1 & \text{if } f_{\lambda_r}(v_{ri}) \geq df_0(v_{ri}), \\ 0 & \text{if } f_{\lambda_r}(v_{ri}) < df_0(v_{ri}), \end{cases}$$

such that $E_{\tilde{\beta}_{ri}} \psi_{ri}^0(Y) = \alpha$, $\tilde{\beta}_{ri} \in \Omega_0$, where v_{ri} is the observed value of V_{ri} and $f_0(v_{ri})$ is the central F probability density. It can also be shown that ψ_{ri}^0 , $i = 1, \dots, k_r$; $r = 2, \dots, p$, maximize

$$\min_{\substack{i=1, \dots, k_r \\ r=2, \dots, p}} \inf_{\tilde{\beta}_{ri} \in \Omega_{1,ri}} E_{\tilde{\beta}_{ri}} \psi_{ri}^0(Y)$$

among all tests ψ_{ri} in $S(\alpha)$, $i = 1, \dots, k_r$; $r = 2, \dots, p$.

The rule ψ_{ri}^0 is equivalent to the following

$$\psi_{ri}^0(y) = \begin{cases} 1 & \text{if } v_{ri} \geq d_{ri}, \\ 0 & \text{if } v_{ri} < d_{ri}, \end{cases}$$

where d_{ri} is determined by

$$P\{V_{ri} \geq d_{ri} | \lambda_{ri} = 0\} = \alpha.$$

Note that the power functions $P\{S_{ri} \geq c_{ri} | \lambda_{ri} \geq \lambda_r\}$ and $P\{V_{ri} \geq d_{ri} | \lambda_{ri} \geq \lambda_r\}$ are increasing in λ_{ri} .

Example:

We use Hald data (Draper and Smith (1981)) to discuss the procedure as follows:

NO	X_1	X_2	X_3	X_4	Y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

The regression model has been established (Draper and Smith (1981)) as

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon_i,$$

$$\epsilon_i \sim N(0, \sigma_0^2), i = 1, 2, \dots, 13.$$

Let

$$p_1 = \min_{2 \leq r \leq p} \min_{1 \leq i \leq k_r} \inf_{\lambda_{ri} \geq \lambda_r} E \psi_{ri}^0(Y) = \min_{2 \leq r \leq p} \min_{1 \leq i \leq k_r} P\{S_{ri} \geq c_{ri} | \lambda_{ri} = \lambda_r\},$$

and

$$p_2 = \min_{2 \leq r \leq p} \min_{1 \leq i \leq k_r} \inf_{\lambda_{ri} \geq \lambda_r} E \psi_{ri}^0(Y) = \min_{2 \leq r \leq p} \min_{1 \leq i \leq k_r} P\{V_{ri} \geq d_{ri} | \lambda_{ri} = \lambda_r\}.$$

Case 1: We use the residual mean square $s^2 = \frac{1}{12} \sum_{i=1}^{13} (Y_i - \bar{Y})^2 = 5.98$ to estimate σ_0^2 as the known value of σ_0^2 . Testing

$$H_{0,ri}: \tilde{\beta} \in \Omega_0 \text{ vs } K_{ri}: \tilde{\beta} \in \Omega_{1,ri}$$

$i = 1, \dots, k_r; r = 2, 3, 4, 5.$

The procedure φ_{ri}^0 is

$$\varphi_{ri}^0(y) = \begin{cases} 1 & \text{if } s_{ri} \geq c_{ri}, \\ 0 & \text{if } s_{ri} < c_{ri}. \end{cases}$$

We list the process to reject inferior models as follows: For $\alpha = 0.05$,

Order	Model	$v_r = 13-r$	$v_r s_{ri}$	$v_r c_{ri}$	Result
1	X_1	11	211.65	19.675	reject
2	X_2	11	151.56	19.675	reject
3	X_3	11	324.31	19.675	reject
4	X_4	11	147.81	19.675	reject
5	$X_1 X_2$	10	9.68	18.307	not reject

Continuing in this fashion, the final decision is: Retain the models $\{X_1, X_2\}, \dots$. Note that the order can be randomly given.

We have obtained $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ as the desired reduced model. The following table is shown the behavior of p_1 when Δ is changed.

$$\alpha = 0.05$$

Δ	1.6	2.0	2.4	2.8
p_1	0.30	0.51	0.69	0.82

It shows that p_1 increases rapidly as Δ increasing.

Case 2: When σ_0^2 unknown.

Testing

$$H_{0,ri}: \tilde{\beta} \in \Omega_0 \text{ vs } K_{ri}: \tilde{\beta} \in \Omega_{1,ri}$$

$i = 1, \dots, k_r; r = 2, \dots, 5.$

The procedure ψ^0 is

$$\psi_{ri}^0(y) = \begin{cases} 1 & \text{if } v_{ri} \geq d_{ri}, \\ 0 & \text{if } v_{ri} < d_{ri}. \end{cases}$$

We list the process to reject inferior models as follows: For $\alpha = 0.1$

Order	Model	v_{ri}	$v_r = 13-4$	d_{ri}	Result
1	X_1	67.88	11	4.07	reject
2	X_2	47.85	11	4.07	reject
3	X_3	107.23	11	4.07	reject
4	X_4	46.60	11	4.07	reject
5	X_1X_2	0.84	10	4.46	not reject

Continuing in this fashion, the final decision is: Retain the models $\{X_1, X_2\}, \dots$. Note that the order can be randomly given.

We also obtain $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ as the desired model. The table of the relation between Δ and p_2 as follows:

$$\alpha = 0.05$$

Δ	1.6	2.0	2.4	2.8
p_2	0.37	0.57	0.74	0.84

Discussion

From the above example, we can see that the procedure φ and ψ are easy to apply to obtain our desired models. We also can guarantee the most conservative power for any model we selected.

Our rule is designed to select all models which are good in the sense of having adequate precision in predicting compared to the true model. When the final selection consists of several models, the experimenter may be guided by practical considerations in choosing one of these. For example, a model with smallest number of variables may be a consideration. In practice, it may be easier to obtain information on some variables than the others; thus a model with variables easy to handle will be preferable. Of course, it may be preferable to build a cost factor in the problem. In the long run, one may randomly choose a model for the selected group.

Acknowledgement

The authors wish to thank Professor Panchapakesan, S. for his helpful critical reading.

References

- [1] Arvensen, J. N. and McCabe, G. P. (1975). Subset selection problems for variances with applications to regression analysis. JASA, 70, 166-170.
- [2] Draper, N. and Smith, H. (1981). Applied regression analysis (2nd ed.) New York: Wiley.
- [3] Hsu, T. A. and Huang, D. Y. (1982). Some sequential selection procedures for good regression models. Comm. Statist. A-Theor. Meth., 11, 411-421.
- [4] Huang, D. Y. and Panchapakesan, S. (1982). On eliminating inferior regression models. Comm. Statist.-Theor. Meth., 11(7), 751-759.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #83-29	2. GOVT ACCESSION NO. AD-A131567	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SELECTION PROCEDURES FOR OPTIMAL SUBSET OF REGRESSION VARIABLES		5. TYPE OF REPORT & PERIOD COVERED Technical
7. AUTHOR(s) Shanti S. Gupta, D. Y. Huang and C. L. Chang		6. PERFORMING ORG. REPORT NUMBER Technical Report #83-29
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE July 1983
		13. NUMBER OF PAGES 10
		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION, DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release, distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Regression variables; optimal subset; selection; testing; chi-square and F distribution; examples.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper deals with selection of an optimal subset of variables in a linear regression model. Based on the criterion of expected residual mean squares, we reject inferior regression models. The derivation of the rule is different from those of the earlier papers in that here we use the simultaneous tests of a family of hypotheses. Using real data, an example is provided to illustrate the application of the proposed procedure.		

END

FILMED

9-83

DTIC